

Designed Coiled Coils Promote Folding of a Recombinant Bacterial Collagen^{*[S]}

Received for publication, December 29, 2010, and in revised form, March 2, 2011. Published, JBC Papers in Press, March 28, 2011, DOI 10.1074/jbc.M110.217364

Ayumi Yoshizumi^{†1}, Jordan M. Fletcher^{S1}, Zhuoxin Yu[‡], Anton V. Persikov[¶], Gail J. Bartlett^S, Aimee L. Boyle^S, Thomas L. Vincent^{S||}, Derek N. Woolfson^{S**2}, and Barbara Brodsky^{‡,††3}

From the [†]Department of Biochemistry, University of Medicine and Dentistry of New Jersey, Robert Wood Johnson Medical School, Piscataway, New Jersey 08854, the ^SSchool of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, United Kingdom, [¶]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544, the ^{||}Bristol Centre for Complexity Sciences, University of Bristol, Bristol BS8 1TR, United Kingdom, the ^{**}School of Biochemistry, University of Bristol, Medical Sciences Building, University Walk, Bristol BS8 1TD, United Kingdom, and the ^{††}Department of Biomedical Engineering, Tufts University, Medford, Massachusetts 02155

Collagen triple helices fold slowly and inefficiently, often requiring adjacent globular domains to assist this process. In the *Streptococcus pyogenes* collagen-like protein Scl2, a V domain predicted to be largely α -helical, occurs N-terminal to the collagen triple helix (CL). Here, we replace this natural trimerization domain with a *de novo* designed, hyperstable, parallel, three-stranded, α -helical coiled coil (CC), either at the N terminus (CC-CL) or the C terminus (CL-CC) of the collagen domain. CD spectra of the constructs are consistent with additivity of independently and fully folded CC and CL domains, and the proteins retain their distinctive thermal stabilities, CL at $\sim 37^\circ\text{C}$ and CC at $>90^\circ\text{C}$. Heating the hybrid proteins to 50°C unfolds CL, leaving CC intact, and upon cooling, the rate of CL refolding is somewhat faster for CL-CC than for CC-CL. A construct with coiled coils on both ends, CC-CL-CC, retains the $\sim 37^\circ\text{C}$ thermal stability for CL but shows less triple helix at low temperature and less denaturation at 50°C . Most strikingly however, in CC-CL-CC, the CL refolds slower than in either CC-CL or CL-CC by almost two orders of magnitude. We propose that a single CC promotes folding of the CL domain via nucleation and in-register growth from one end, whereas initiation and growth from both ends in CC-CL-CC results in mismatched registers that frustrate folding. Bioinformatics analysis of natural collagens lends support to this because, where present, there is generally only one coiled-coil domain close to the triple helix, and it is nearly always N-terminal to the collagen repeat.

The collagen triple helix and the α -helical coiled coil (CC)⁴ are well characterized superhelical motifs in proteins (1–3). They form rod-like structures, which are directed by clear

amino acid patterns in their sequences. Collagen triple helices require glycine as every third residue and often have a high imino acid (proline and hydroxyproline) content. These features lead to the formation of polyproline II helices, which trimerize via interchain hydrogen bonding and close packing to form the collagen triple helix (see Fig. 1A). By contrast, most coiled coils are low in glycine and proline and have a so-called heptad, or related repeats in which hydrophobic residues alternate three and four residues apart. This pattern promotes the formation of amphipathic α -helices that combine via their hydrophobic faces to form rope-like helical bundles (Fig. 1, B and C).

These supercoiled collagen and α -helical coiled-coil structures were first elucidated as the major elements in fibrous proteins but have since been found in a wide range of proteins, including globular and membrane-spanning structures (2, 3). Some proteins contain both collagen and α -helical coiled-coil domains. For instance, three-stranded coiled coils occur immediately C-terminal to collagen triple helices in lung surfactant apoprotein D, lung surfactant apoprotein A, mannose-binding protein, and other collectins (4), whereas the macrophage scavenger receptor has a coiled coil N-terminal to a collagen triple helix (5). Previous sequence analyses suggest that putative coiled-coil domains are often found in members of the collagen superfamily, and these coiled coils may serve as oligomerization domains important for collagen assembly (6). *In vitro* triple helix formation from three (Gly-Xaa-Yaa)_n polypeptide chains is an inherently slow process (7), and it appears that adjacent coiled coils facilitate proper registration, nucleation, and folding of the triple helix. Here, we explore the relationship between coiled-coil domains and an adjacent collagen triple helix using a designed coiled coil and a recombinant bacterial collagen.

Although collagens were originally thought to be restricted to multicellular animals, a number of collagen-like triple helix domains have been identified in bacteria (8). Of these, the cell-surface proteins from *Streptococcus pyogenes* are among the best characterized in terms of structure and function. Scl2 (*S. pyogenes* collagen-like protein 2) contains an N-terminal globular domain (denoted V), a (Gly-Xaa-Yaa)₇₉ collagen-like domain (CL), and linker and transmembrane regions (9). A recombinant construct with the globular V domain and the

^{*} This work was supported, in whole or in part, by National Institutes of Health Grant GM60048 (to B. B.). This work was also supported by Biotechnology and Biological Sciences Research Council Grant BB/G008833/1 (to D. N. W.).

^[S] The on-line version of this article (available at <http://www.jbc.org/>) contains supplemental Figs. S1–S3 and additional references.

¹ Both authors contributed equally to this work.

² To whom correspondence may be addressed. E-mail: D.N.Woolfson@bristol.ac.uk.

³ To whom correspondence may be addressed. E-mail: Barbara.Brodsky@tufts.edu.

⁴ The abbreviations used are: CC, coiled coil; CL, collagen; MRE, mean residue ellipticity; Fmoc, N-(9-fluorenyl)methoxycarbonyl; deg, degrees.

adjacent triple helix, designated as V-CL, has been expressed in *Escherichia coli* and shown to form a triple helix structure with a conformation and thermal stability similar to that of mammalian collagens (see Fig. 2A) (9–11).

The N-terminal V domain has been shown to be a trimerization domain and is essential for *in vitro* refolding of the CL triple helix (9, 12). In addition, the *S. pyogenes* V domain can assist correct folding of a heterologous triple helix sequence from *Clostridium perfringens*, which is incapable of folding in its original context (12, 13). Although a three-dimensional structure of the V domain has yet to be determined, two α -helical regions are predicted from its sequence. The isolated V domain has an α -helical CD spectrum (9, 12), suggesting that the Scl2 bacterial collagen protein may represent another instance where an adjacent α -helical domain promotes triple helix formation. The two α -helical regions in the V domain were assigned previously as coiled coils (9, 12), but re-examination using the currently available prediction tools indicates that the coiled-coil nature of regions are predicted with ~25% confidence or less (supplemental Fig. S1).

For the present study, we constructed recombinant fusion proteins in which the natural N-terminal V domain of the Scl2 protein was replaced by a parallel, homotrimeric coiled coil of *de novo* design and high thermal stability. Specifically, a four-heptad sequence was placed either N- or C-terminal to, or at both ends of, the collagen domain. The effect of the position of the coiled-coil domains on folding, stability, and assembly of the triple helix was investigated by equilibrium, thermal unfolding, and kinetic CD spectroscopy experiments. These data indicate that each domain retained its independent stability in all cases, largely unaffected by the presence of the other domain in the same protein. Conditions were established where the collagen domain unfolded, whereas the coiled coil remained structured. In these cases, the refolding kinetics for the collagen triple helix depended markedly on the position of the coiled-coil domain. Intriguingly, the construct with coiled-coil domains at both ends of the collagen domain folded more slowly and less completely than those with a single coiled coil, which suggests that some structural frustration occurs if collagen triple helices are nucleated from both ends. The biological consequences of this hypothesis are explored through bioinformatics analysis of protein sequences that harbor both collagen and coiled-coil repeats.

EXPERIMENTAL PROCEDURES

Peptide Synthesis—Rink amide ChemMatrixTM resin was obtained from PCAS Biomatrix, Inc. (St.-Jean-sur-Richelieu, Canada); Fmoc-L-amino acids were obtained from AGTC Bio-products (Hessle, UK); 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate was obtained from GL Biochem (Shanghai, China); all other reagents were of peptide synthesis grade and obtained from Thermo Fisher Scientific (Loughborough, UK). Peptide BS-3pCC4 was synthesized on a 0.1-mmol scale on Rink amide resin using a LibertyTM microwave peptide synthesizer (CEM; Matthews, NC) employing Fmoc solid-phase techniques (for review, see Ref. 14) and systematically repeated steps of coupling and deprotection interspaced with washings (5 × 7 ml dimethylformamide).

Coupling was performed as follows: Fmoc amino acid (5 eq), 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (4.5 eq), and diisopropylethylamine (10 eq) in dimethylformamide (7 ml) for 5 min with 20-watt microwave irradiation at 75 °C. Deprotection was performed as follows: 20% piperidine in dimethylformamide for 5 min with 20-watt microwave irradiation at 75 °C. Following linear assembly, the peptide was acetylated (acetic anhydride (3 eq) and diisopropylethylamine (4.5 eq) in dimethylformamide (7 ml) for 20 min) and then cleaved from the resin with concomitant removal of side chain-protecting groups by treatment with a cleavage mixture (10 ml) consisting of TFA (95%), triisopropylsilane (2.5%), and H₂O (2.5%) for 3 h at room temperature. Suspended resin was removed by filtration, and the peptide was precipitated in ice-cold diethyl ether, centrifuged, and then the pellet was dissolved in 1:1 MeCN/H₂O and freeze-dried. Purification was performed by RP-HPLC using a Kromatek C18 reverse phase column (semi-micro, 5 μ m, 100 Å, 10 mm inner diameter × 150 mm long). Eluents used were as follows: 0.1% TFA in H₂O (A) and 0.1% TFA in MeCN (B). The peptide was eluted by applying a linear gradient (at 3 ml/min) of 20% to 80% B over 40 min. Fractions collected were examined by MALDI-TOF mass spectrometry, and those found to contain exclusively the desired product were pooled and lyophilized. Analysis of the purified final product by RP-HPLC indicated a purity of >95%. Successful synthesis was confirmed by MALDI-TOF mass spectrometry (BS-3pCC4: m/z [M + H]⁺_{observed}: 3466.06; m/z [M + H]⁺_{theoretical}: 3465.95).

Construction of pCold II-CC-CL, pCold II-CL-CC, and pCold II-CC-CL-CC and Protein Expression—A DNA fragment encoding the BS-3pCC4 peptide sequence was synthesized by GenScript and inserted in place of the coding region for the variable globular domains of pCold III-V-CL and pCold III-CL-V (10, 11, 15) bearing a His₆ tag at the N terminus. For pCold II-CC-CL-CC, the coiled-coil gene sequence was inserted at both the 5' and 3' ends of the coding regions for CL in pCold II-CL. All resulting plasmids were confirmed by DNA sequencing and then transformed into *E. coli* BL21 strains. Cells were cultured in 5 ml of M9 casamino acid medium containing ampicillin (50 μ g/ml) and incubated at 37 °C for 12 h. The cultures were then transferred into 250 ml of M9 casamino acid medium containing ampicillin (50 μ g/ml) and incubated at 37 °C for approximately 4 h until the A₆₀₀ value reached 0.8 absorbance units. The culture was shifted to room temperature, and 1 mM isopropyl 1-thio- β -D-galactopyranoside was added to induce protein expression. After overnight expression, cells were harvested by centrifugation and disrupted by French pressing. Cellular debris was removed by centrifugation at 4 °C. All proteins were found in the soluble supernatant fraction. The pure proteins were obtained by elution with imidazole from a nickel-Sepharose resin column initially equilibrated with binding buffer (20 mM phosphate buffer, pH 7.4, 500 mM NaCl, 20 mM imidazole) at room temperature, as described previously (11).

CD Spectroscopy—CD spectra of BS-3pCC4 were recorded using a JASCO J815 spectropolarimeter, whereas recombinant collagen/coiled coil constructs were examined using an AVIV Model 62DS spectropolarimeter (Aviv Associates, Inc., Lake-

Fusing Coiled-coil and Collagen Domains

wood, NJ). Peptide concentrations were determined by UV absorption at 280 nm ($\epsilon(\text{Trp}) = 5690 \text{ mol}^{-1} \text{ cm}^{-1}$; $\epsilon(\text{Tyr}) = 1280 \text{ mol}^{-1} \text{ cm}^{-1}$). Ellipticities (millidegrees) were converted to mean residue ellipticity (MRE; $\text{deg cm}^2 \cdot \text{dmol}^{-1}$) by normalizing for amide bond concentration and cuvette path length. In the absence of aromatic residues, the concentration of the collagen domain (CL) was estimated by weighing and by absorbance at 214 nm and also by subtracting the CD spectrum of the isolated V domain from the CD spectrum of the intact V-CL molecule, giving $\text{MRE}_{222} \sim 7000 \text{ deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}$ and $\text{MRE}_{198} \sim -60,000 \text{ deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}$ for CL.

Secondary Structure and Thermal Stability—The four-heptad trimeric coiled coil, BS-3pCC4, the collagen triple helix domain (CL), the coiled coil/collagen constructs (CC-CL and CL-CC), and CC-CL-CC were prepared for analysis at 20 μM in PBS at pH 7.0. Complete spectra were obtained at 0 °C and 50 °C from 260 nm to 196 nm recording points at every 0.5 nm for 4 s using a bandwidth of 1 nm, averaging three scans for each sample. Spectra obtained for BS-3pCC4 and CL were also used as a basis for determining theoretical spectra for the coiled coil/collagen constructs to which the experimentally observed spectra were compared. Theoretical spectra were calculated by taking into account the relative sizes of both the BS-3pCC4 (32 residues) and CL (238 residues) domains, e.g. $\text{MRE}(\text{CC-CL})_{\text{theoretical}} = (32/(238 + 32)) \times \text{MRE}(\text{BS-3pCC4}) + (238/(238 + 32)) \times \text{MRE}(\text{CL})$.

Experiments examining the thermal denaturation of coiled coil/collagen constructs were also performed at 20 μM concentration in PBS (pH 7.0). Molar ellipticity was recorded at 220 nm from 0 to 100 °C at an average ramp rate of 0.1 °C/min. The T_m was determined as the temperature at which the fraction folded was equal to 0.5 in the curve fitted to the trimer-to-monomer transition. Owing to the high thermal stability of the coiled-coil component of the coiled coil/collagen constructs, the thermal denaturation curves of BS-3pCC4 and CC-CL were also examined in the presence of guanidine HCl (0, 1, 2, and 3 M in PBS, pH 7.0). Thermal melts for CC-CL were performed by measuring MRE at 220 nm from 0 to 100 °C at 0.1 °C/min, whereas the much faster folding coiled-coil peptide BS-3pCC4 was examined from 5 °C to 90 °C at a rate of 0.67 °C/min.

Kinetic Analysis of Coiled Coil/Collagen Refolding—V-CL and fused coiled-coil and collagen constructs (20 μM) were denatured at 50 °C for 30 min in PBS (pH 7.0), a temperature that is sufficient to denature the collagen triple helix but leaves the coiled-coil domain intact. After denaturation, the sample was immediately transferred to CD cells pre-equilibrated at 0 or 25 °C, and the ellipticity at 220 nm was monitored as a function of time (time constant, 2 s; time interval, 10 s) to follow refolding of the collagen triple helix. Because the folding curves did not fit any simple kinetics, the half-time of refolding was defined as the time for the fraction folded to reach 50% of the original $\text{MRE}_{220 \text{ nm}}$ value at low temperature.

Analytical Ultracentrifugation—Sedimentation equilibrium experiments for the BS-3pCC4 peptide were conducted in a Beckman-Optima XL-I analytical ultracentrifuge at 20 °C using an An-60 Ti rotor. Solutions were prepared in PBS (pH 7.4) at peptide concentrations in the range of 25–400 μM . Centrifugation speeds were between 23,000–54,000 rpm. Data sets were

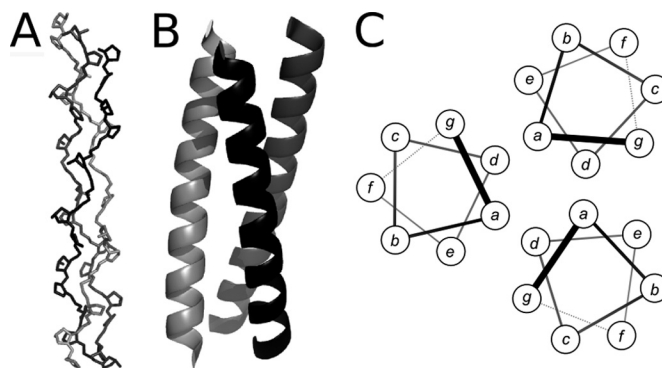


FIGURE 1. Collagen and α -helical coiled-coil structures. A, backbone structure of a model peptide that adopts a collagen-like triple helix (Protein Data Bank code 2DRT). B, ribbon representation of a designed trimeric α -helical coiled-coil peptide (Protein Data Bank code 1GCM). C, helical wheels for the latter showing the positions and juxtapositions of residues in the coiled-coil heptad repeat **abcdefg**.

fitted to a single, ideal species model using Ultrascan (16). The partial specific volumes (0.7716 ml/g) of the peptide and the solvent density (1.0054 g/ml) were calculated using Sednterp (17).

Bioinformatics—The list of representative architectures for the protein family PF01391 (collagen triple helix repeat) was retrieved from the Pfam database (18). Positions of coiled-coil domains and collagen repeats were identified from these sequences and augmented with data from Marcoil (19), a hidden Markov model-based coiled-coil sequence predictor, using a confidence level of 0.9. For those sequences where the coiled-coil domain and the collagen repeats were within 15 residues, oligomer state predictions were carried out using a combination of SCORER (20) and LOGICOIL.⁵ Given a coiled-coil sequence, both algorithms predict whether that sequence forms a dimer or a trimer, based on evidence garnered from training sets of sequences of known oligomeric state. From these, conservative definitions were taken; if SCORER and LOGICOIL concurred in their oligomer-state prediction, that state was assigned; if they disagreed, no state was assigned. Two of the sequences had structurally characterized coiled-coil domains identified from Pfam and did not require oligomeric state prediction.

RESULTS

Design and Characterization of Synthetic α -Helical Coiled Coil, BS-3pCC4—A short peptide sequence predicted to form a highly stable parallel trimeric coiled coil was designed to be amenable to characterization via solution phase biophysical methods, as well as subsequent gene synthesis and cloning into Scl2 constructs for gene expression and protein production. The starting point for the design was the coiled-coil heptad repeat, designated **abcdefg** and visualized on the helical wheel of Fig. 1C. Different amino acid combinations at the predominantly hydrophobic, **a** and **d** positions direct different oligomer states (21, 22). For the required design, the combination of **a** = **d** = Ile was used to prescribe parallel trimers. The interface was cemented further with oppositely charged residues at

⁵ T. L. Vincent, P. J. Green, and D. N. Woolfson, unpublished results.

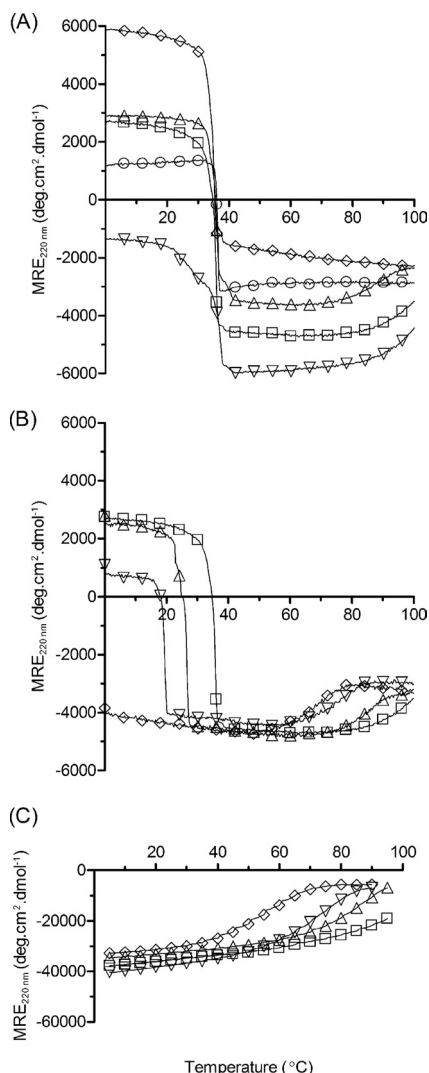


FIGURE 4. Thermal unfolding behavior of the coiled-coil/collagen constructs. A, V-CL (○), CL (◇), CC-CL (□), CL-CC (△), and CC-CL-CC (▽) at 20 μ M peptide/protein in PBS, pH 7.0. CC-CL (B) and BS-3pCC4, *i.e.* CC alone (○), in the presence of guanidine HCl at 0 M (□), 1.0 M (△), 2.0 M (▽), and 3.0 M (◇). T_m of coiled coil denaturation in 3 M guanidine HCl, CC-CL (69 °C, B (◇)) and BS-3pCC4 (55 °C, C (◇)).

were transformed into *E. coli* strain BL21, and the constructs were expressed at room temperature, which was established previously as optimal for expression (11). All constructs expressed as soluble proteins and were purified on a nickel-Sepharose column. The eluted proteins were detected as single bands on SDS-PAGE near the expected molecular weight position, confirming their identity and purity.

Secondary Structure and Thermal Stability of Fused Coiled-coil and Collagen Constructs—The conformations of the coiled-coil peptide, BS-3pCC4, the isolated CL collagen triple helix domain, and the fusion proteins were compared by CD spectroscopy. As described above, BS-3pCC4 showed a typical α -helical spectrum (Fig. 3A). The CD spectrum of recombinant CL domain alone (expressed as His₆-CL) had the typical triple helix maximum at 220 nm and a minimum at 198 nm as expected for a collagen-like triple helix, with a ratio of the positive to negative peaks close to the 0.12 value expected for a fully triple helical molecule (Fig. 3A) (24).

The recombinant CC-CL and CL-CC proteins both gave CD spectra at 0 °C with a maximum at 220 and a minimum at 198 nm as well, but the magnitudes of these peaks were lower than for CL alone. From the sequences of these constructs, the four-heptad coiled coil and collagen (Gly-Xaa-Yaa)₇₉ domains account for 10 and 90% of the residues, respectively. On the basis of these results, we calculated predicted spectra for the two hybrid sequences, *i.e.* using the observed CD spectra for CC domain (0.1 fraction) and of the CL domain (0.9 fraction). These calculated spectra were very similar to those observed (Fig. 3, C–F). This excellent additivity indicates that the coiled-coil and collagen triple helix domains retain their original structures without perturbation when in the same molecule. The low temperature CD spectrum of the larger construct, CC-CL-CC (Fig. 3G), had a reduced MRE₂₂₀ of $-1237 \text{ deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$, which is consistent with the additional coiled-coil domain contributing negative ellipticity at this wavelength. The observed and calculated spectra for this construct are in general agreement; although there is a larger discrepancy than found for the single coiled-coil constructs.

Monitoring the MRE₂₂₀ with increasing temperature showed that all three fusion proteins had sharp thermal unfolding transitions centered ~ 36 – 37 °C (Fig. 4A). Interestingly, this T_m value is indistinguishable from that observed for V-CL and CL proteins. Because the BS-3pCC4 peptide does not unfold until very high temperatures, these melting transitions likely reflect denaturation of only the collagen triple helix. To confirm this, CD spectra for BS-3pCC4, CC-CL, CL-CC, and CC-CL-CC were recorded at 50 °C, (Fig. 3, right panels) where the collagen triple helix portion is expected to be unfolded, whereas the coiled-coil region is expected to be intact. The observed CD spectra for CC-CL and CL-CC showed typical α -helix minima at 208 and 222 nm, with magnitudes in excellent agreement with those calculated from a weighted average of spectra for isolated folded BS-3pCC4 and unfolded CL at 50 °C (Fig. 3B). The observed CD spectrum for CC-CL-CC at 50 °C showed more pronounced minima at 208 and 222 nm, consistent with the expected increased α -helix content. However, in this case, the observed magnitudes were less than calculated. It is possible that there is incomplete unfolding of the central triple helix domain in this highly constrained context, or that the coiled-coil and collagen domains perturb each other in some other way.

The very high stability of the coiled coil made it difficult to observe directly the melting of both domains of the hybrid molecule. Therefore, increasing molarities of GdnHCl were introduced to lower the stability of the coiled coil (Fig. 4C). Addition of 1–3 M GdnHCl allowed observation of the independent thermal melting of the collagen triple helix domain (decreasing MRE₂₂₀), followed by melting of the α -helical coiled coil at higher temperatures (increasing MRE₂₂₀) (Fig. 4B). Clearly, there is a region of intermediate temperature where the triple helix is fully melted, whereas the coiled-coil structure remains intact. This facilitated the following refolding experiments.

Refolding Kinetics of Fused Coiled-coil and Collagen Domains—Previous studies show that recombinant V-CL refolds *in vitro* following heat denaturation, whereas the collagen triple helix domain CL alone does not refold significantly even after weeks

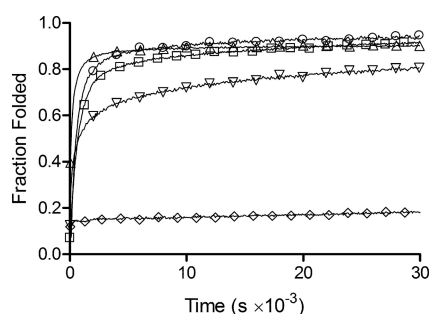


FIGURE 5. **Refolding of coiled-coil/collagen constructs.** The collagen component of all constructs was denatured by heating to 50 °C for 30 min, samples were then returned to 0 °C and refolding monitored by CD spectroscopy at 220 nm: V-CL (○), CL (◇), CC-CL (□), CL-CC (△), and CC-CL-CC (▽).

TABLE 1

Summary of biophysical characterization of the coiled-coil and collagen constructs

All data were collected at peptide/protein concentrations of 20 μ M in PBS, pH 7.0. ND, not determined.

| Construct | MRE _{220 nm} ^a | MRE _{198 nm} ^a | <i>T_m</i> ^b | <i>t</i> _{1/2} ^c |
|-----------|--|--|-----------------------------------|--------------------------------------|
| | deg cm ² dmol ⁻¹ | deg cm ² dmol ⁻¹ | °C | min |
| BS-3pCC4 | -30,674 | 28,437 | ND | ND |
| V-CL | 1320 | -7373 | 36.2 | 8.85 |
| CC-CL | 2797 | -29,500 | 35.3 | 11.8 |
| CL-CC | 3867 | -47,972 | 35.8 | 2.02 |
| CC-CL-CC | -1232 | -20,917 | 35.5 | 219.5 |
| CL-V | 1470 | -53,162 | 36.0 | 21.2 |
| CL | 7134 | -60,503 | 34.7 | >1000 |

^a MRE at 220 nm and 198 nm from CD studies conducted at 0 °C presented in Fig. 3.

^b Melting temperature of the collagen component of the fusion from the thermal denaturation experiments presented in Fig. 4. The melting temperature of the coiled-coil BS-3pCC4 peptide was >90 °C, and a precise *T_m* value could not be determined.

^c Half-times for refolding at 0 °C of the CL domains in each construct following a 30-min thermal denaturation at 50 °C as presented in Fig. 5.

at low temperature (Fig. 5) (9, 11, 12). The refolding of the collagen triple helix within the coiled-coil fusions described herein was monitored by following the recovery of MRE₂₂₀ at 0 °C after heating to 50 °C for 30 min (PBS, pH 7.0), *i.e.* where the collagen triple helix is initially fully denatured, but the coiled coil is preserved throughout the experiment.

The constructs containing a coiled-coil domain on either end showed substantial refolding of the collagen triple helix, reaching a final value close to 100% after 5 days (Fig. 5). The overall folding kinetics and extent of folding was generally similar to that seen for V-CL, but the folding rates at early time points differed between constructs (Fig. 5). In particular, the folding of CL-CC was faster than CC-CL. This contrasts with the constructs harboring the V domain, where V-CL, the natural arrangement, refolds faster than the permuted CL-V (Table 1) (12). Notably, the protein with coiled-coil domains at both ends of the collagen triple helix, CC-CL-CC, showed a markedly slower folding rate and lower percent recovery than CC-CL and CL-CC (Fig. 5). Refolding of CC-CL-CC was also studied at a higher temperature, 25 °C, where misfolded states are more likely to unfold and refold properly. However, the refolding at 25 °C resulted in much slower rates and poorer refolding efficiencies than at 0 °C for CC-CL-CC (supplemental Fig. S3).

DISCUSSION

The sequences of the recombinant proteins characterized here contain four-heptad segments designed to form a stable

parallel, three-stranded coiled coil adjacent to a proline-rich (Gly-Xaa-Yaa)₇₉ domain that forms a stable collagen-like triple helix. As we discuss further below, there are natural precedents for such arrangements. Therefore, the juxtaposition of these domains in relatively “pure” forms within the same polypeptide makes an interesting study, particularly given the structural similarities and differences between the two motifs and the likely different mechanisms of folding.

Both the collagen triple helix and the three-stranded coiled coil are rod-like assemblies in which three chains adopt defined secondary structures, combine, and intertwine to form supercoiled structures. However, this is where the similarities end. The collagen triple helix comprises three polyproline II-like helical chains supercoiled about a common axis, with hydrogen bonding between adjacent chains. The chains are staggered by one residue to allow close packing of the invariant glycine residues within the core of the triple helix (Fig. 1A). In contrast, in three-stranded α -helical coiled coils, each chain folds as a self-contained, internally hydrogen-bonded α -helix. Patterns of hydrophobic and polar residues on the surface of these structures drive helix association, which is usually perfectly in register (Fig. 1B).

These sequence and structural differences affect the mechanism of folding of the two domains. *In vitro* assembly of (Gly-Xaa-Yaa)_n repeats into triple helices is slow, with folding taking from minutes to days for collagens and model triple helical peptides (7, 25). Factors that limit triple helix folding include the following: 1) slow *cis-trans* isomerization of imino acids, which can limit nucleation and propagation of the secondary structure; 2) ternary association; 3) (related to 1) the need for chains to adopt correct dihedral angles to allow close packing and hydrogen bonding between chains; and 4) the possibility of misfolding and misalignment of chains due to the repetitiveness of the sequences (25, 26). By contrast, the more modular assembly of α -helical coiled coils makes their folding generally rapid and efficient on the μ s-ms time frame (27). However, we note for α -helical coil coils that some require particular sequence motifs to trigger folding and assembly (28) and that the more general structure and mechanism of assembly can lead to promiscuity (21, 23) as many more coiled-coil architectures are possible in addition to the three-stranded parallel type (3, 29).

The study presented here for constructs containing both the trimeric α -helical coiled-coil and the collagen triple helix motif suggests little influence of one domain on the conformation of the other, which may not be surprising for these rod-like, linear motifs. At least in aqueous buffers without denaturants (see below), this independence extended to the thermal stabilities of the two domains; the \sim 37 °C thermal transition for the collagen triple helix remained unchanged regardless of the location of the coiled-coil domain(s), and the thermal stability of the coiled coil remained high and similar to that for an isolated, synthetic peptide of the same sequence. The use of a highly stable coiled coil allowed melting of the collagen triple helix domain, whereas the coiled-coil domain(s) remained intact, *i.e.* above 40 °C. This contrasts with the melting of a construct based on the full-length natural protein, V-CL, where the two domains are closer in thermal stability (*T_m* of the V domain \sim 46 °C (12)

Fusing Coiled-coil and Collagen Domains

and that of CL $\sim 36 - 37^\circ\text{C}$ (Fig. 4A)), and the domains influence each other. The V-CL fusion protein melts at $\sim 37^\circ\text{C}$, which is considerably lower than the isolated V domain. This indicates coupling in V-CL, whereby the V domain starts unfolding when the CL domain unfolds.

Replacement of the more complex natural V domain by the far more stable α -helical coiled-coil sequence changed the nature of thermal transitions in the system. Thermal unfolding of the stable CC domain was not observed in the recombinant constructs until 3 M guanidine HCl was added (Fig. 4B). At this point, the CL domain appears to be completely unfolded from the outset of the thermal denaturation experiments (*i.e.* 0°C), and the CC domain unfolds with a sigmoidal curve with a T_m of 69°C . This compares with 55°C for the BS-3pCC4 peptide alone. Thus, the construct stabilizes the CC domain in some way. We suggest that any residual structure in or near the junction of the two domains (*i.e.* some remaining collagen or coiled-coil triple helix) will stabilize this domain through an entropic effect. Otherwise, we have no explanation for this effect at present, and further high resolution structural studies will be required to resolve this issue.

Turning to the kinetics of refolding and assembly upon cooling from 50°C , the CL domain alone assembles slowly and incompletely over a period of weeks, whereas all of the engineered constructs combining the CC or V domains with a CL unit fold faster and to near completion. The ability of a CC domain on either the N or C terminus to promote triple helix formation in the CL domain is consistent with previous studies, showing that triple helix formation can be nucleated from either end (26). However, in contrast to the equilibrium CD spectra and thermal melting curves, the refolding rates do depend on the location of the CC domain. In the natural protein, the V domain is N-terminal to the CL domain, and moving it to the C-terminal slows down folding. For the designed fusions, however, the reverse is true; a C-terminal CC domain is more effective in promoting folding. Moreover, the V-CL protein is completely denatured at 50°C , with no remaining α -helical structure; yet this protein refolds as quickly as the engineered CC-CL fusion where the coiled-coil structure is intact, and only the triple helix portion refolds. Very rapid refolding; *i.e.* within the 1 to 2 min it takes to cool the sample from 50 to 0°C , is observed for the isolated V domain,⁷ consistent with the V domain undergoing rapid trimerization prior to CL folding in V-CL. Taken together, the data support a relation between the V and CL domains that is highly evolved in the natural protein, such that the N-terminal original placement is optimized for folding, whereas the C-terminal is not. Such a relationship would most likely be disrupted when the V domain is placed C-terminal to CL. In our engineered constructs, the coupling between the domains seen in the natural protein appears to be lost. This could be related to the very large difference in stability between the CL and CC domains, compared with the closer T_m values of CL and V, or alternatively, to the inclusion of flexible Gly linkers between the CC and CL domains. Thus, in our engineered constructs, we propose that it is simply the proximity of

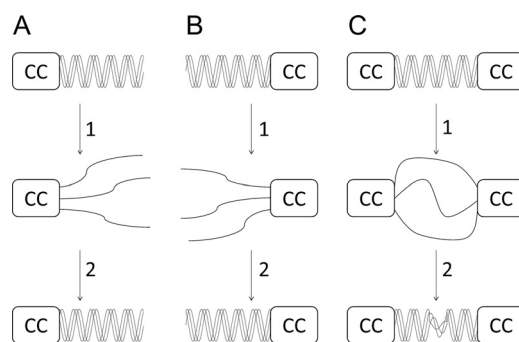


FIGURE 6. Scheme of the possible mechanisms of collagen refolding in the three CL/CC constructs CC-CL (A), CL-CC (B), and CC-CL-CC (C). In each case, step 1 depicts the CL domains undergoing thermal denaturation, and step 2 depicts their refolding. The constructs with a single CC domain (A and B) refold completely, whereas that with a CC at both ends folds incompletely. We posit that this is due to register mismatch when collagen assembly is nucleated from both ends.

the CC domains that promote CL assembly from either end rather than some specific interaction. A full understanding of this awaits structural resolutions of the V domain and the various V/CC-CL constructs, which are currently underway.

Incorporation of CC domains at both ends of the CL domain did not affect CL stability, although the triple helix content appeared a little lower than expected. However, this construct, CC-CL-CC, refolded from 50°C with a half-time ~ 100 times longer than that of the fastest folding construct, CL-CC (Table 1). Thus, it appears that bracketing the collagen triple helix with two coiled coils hinders the assembly of the former. We posit that in the single coiled-coil constructs, folding of the collagen domain is initiated from the end proximal to the folded coiled coil and propagates toward the distal end. In this way, any misfolding or out-of-register alignment of the polypeptide chains is either avoided or readily corrected by unwinding. However, with the chains stabilized by trimeric coiled coils at both ends, folding and propagation can be initiated from both ends with the possibility of mismatched registers near the middle of the CL domain (Fig. 6). The results on the recombinant CC-CL-CC contrast with studies on bovine type III collagen, where the folding rate and folding directionality of the molecule with disulfide bonds at both ends is very similar to that with a disulfide bond only at the C terminus (26). The longer length of the type III collagen and the discontinuity between the N-terminal disulfide and the main triple helix may partially explain this difference. Examination of a simple model of three twisted ropes tethered at both ends revealed difficulty in unfolding and unwinding the central triple helix, whereas both ends are kept fixed, an observation that may be more relevant for the shorter CL triple helix.

Our data are consistent with and indeed shed further light on the domain structure of natural collagen proteins. In these, the presence of non-collagen domains, such as the V domain in Scl2, adjacent to the tripeptide repeats leads to rapid trimerization at one end and forces the three chains of the (Gly-Xaa-Yaa)_n sequence into close proximity at that end, promoting proper registration and nucleation of the triple helix. As suggested by Hoppe *et al.* (4), such results show that the non-collagen protein sequences, which have no inherent staggered arrangement of chains are sufficient to promote triple helix

⁷ E. Hwang, Z. Yu, and B. Brodsky, unpublished data.

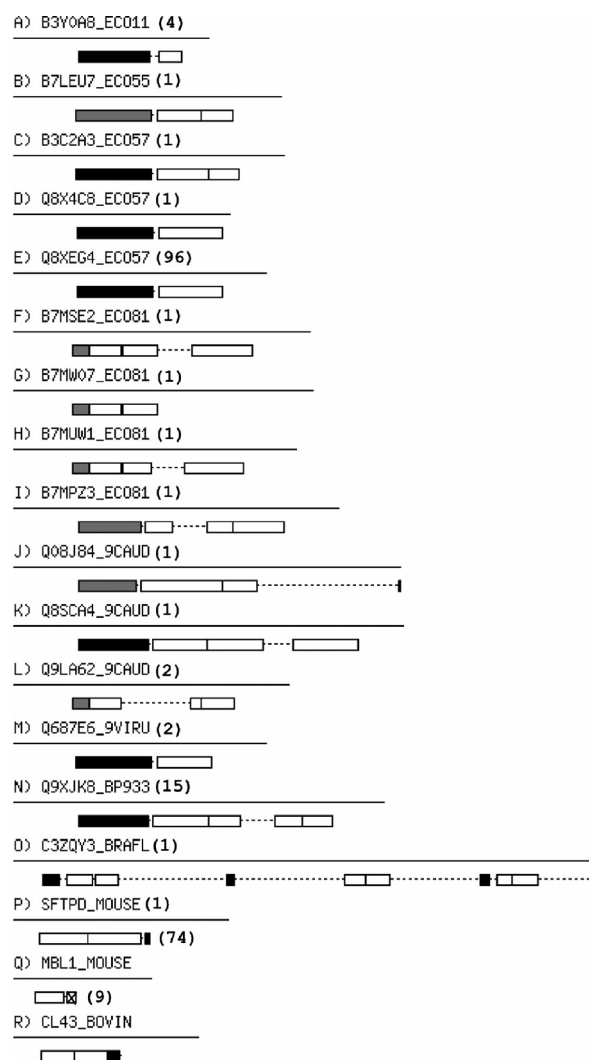


FIGURE 7. Architectures of unique proteins that contain collagen and α -helical coiled-coil domains within 15 residues of each another. Each architecture is represented by one example protein sequence, shown as a line and with domains blocked as follows: white, collagen repeats; black, predicted trimeric coiled coils; cross-hatched, predicted dimeric coiled coils; and gray, those predicted as coiled coil, but with inconclusive oligomer state. Numbers of protein sequences in Pfam for each architecture are shown in parentheses after the example sequence identifier. Oligomeric states could be unambiguously assigned in the case of P and R, which had structurally characterized coiled-coil domains identified in Pfam. Image created using the Perl Bio::Graphics module.

nucleation in the correct one-residue stagger. In an extensive study (6), which examined the distribution of collagen and coiled coil domains, McAlinden *et al.* (6) state that “coiled-coil domains are present in most members of the collagen superfamily, located either before, after or between collagen-like regions, suggesting a general role in triple helix assembly.” We sought to investigate this further with a focus, consistent with our study, on protein architectures containing coiled-coil and collagen domains proximal (*i.e.* <15 residues) to one another, as opposed to the more general cases examined by McAlinden *et al.* (6).

We used the Pfam database (18) to identify proteins that contain a collagen domain. From these, we took a data set of 392 proteins with unique domain architectures, that is, different distributions of collagen repeats and other domains along the

protein sequence. Marcoil (19) was then used to predict all coiled-coil domains within these sequences. This returned 54 protein architectures that contained a coiled-coil region along with the collagen repeat. Only representative proteins with a coiled-coil domain within 15 residues of a collagen repeat are highlighted in Fig. 7. Two things are immediately apparent from this figure: first, all but two of these architectures have coiled-coil regions predicted on just one side of the collagen domain, fully consistent with our data and hypothesis; second, in the vast majority of cases, all except three sequences, the coiled-coil regions are N-terminal to the collagen region. Using the coiled-coil oligomer-state prediction algorithms SCORER (20) and LOGICOIL,⁵ 8/15 of the N-terminal coiled coils predicted as trimers, and 7/15 predictions were inconclusive. Two of the three C-terminal coiled coil domains have been structurally characterized and are trimeric; the third is predicted to be a weak dimer. The absence of any collagen domains flanked by coiled coils on both ends is consistent with the poorer folding that we observe with such an arrangement experimentally.

Perhaps the most important observation from the bioinformatics analysis, however is that when coiled-coil and collagen regions are found within 15 residues of one another, they are almost exclusively arranged such that the coiled coil is N-terminal to the collagen domain; this extended to the set of 54 architectures, in which only six have coiled-coil domains C-terminal of a collagen domain. These results largely concur with those reported by McAlinden *et al.* (6) but highlight that when only proteins containing adjacent (<15 residues apart) coiled-coil and collagen domains are considered (as opposed to a more general case), there is a clear preference for architectures that place trimeric coiled coils N-terminal to their respective collagen domain. This suggests a possible role for coiled coils in the folding and assembly of collagen domains as they emerge from the protein synthesis machinery in the cell.

Acknowledgment—We thank Eileen Hwang for helpful discussions.

REFERENCES

1. Beck, K., and Brodsky, B. (1998) *J. Struct. Biol.* **122**, 17–29
2. Brodsky, B., and Persikov, A. V. (2005) *Adv. Protein Chem.* **70**, 301–339
3. Lupas, A. N., and Gruber, M. (2005) *Adv. Protein Chem.* **70**, 37–78
4. Hoppe, H. J., Barlow, P. N., and Reid, K. B. (1994) *FEBS Lett.* **344**, 191–195
5. Kodama, T., Freeman, M., Rohrer, L., Zabrecky, J., Matsudaira, P., and Krieger, M. (1990) *Nature* **343**, 531–535
6. McAlinden, A., Smith, T. A., Sandell, L. J., Fichoux, D., Parry, D. A., and Hulmes, D. J. (2003) *J. Biol. Chem.* **278**, 42200–42207
7. Baum, J., and Brodsky, B. (1999) *Curr. Opin. Struct. Biol.* **9**, 122–128
8. Rasmussen, M., Jacobsson, M., and Björck, L. (2003) *J. Biol. Chem.* **278**, 32313–32316
9. Xu, Y., Keene, D. R., Bujnicki, J. M., Höök, M., and Lukomski, S. (2002) *J. Biol. Chem.* **277**, 27312–27318
10. Mohs, A., Silva, T., Yoshida, T., Amin, R., Lukomski, S., Inouye, M., and Brodsky, B. (2007) *J. Biol. Chem.* **282**, 29757–29765
11. Yoshizumi, A., Yu, Z., Silva, T., Thiagarajan, G., Ramshaw, J. A., Inouye, M., and Brodsky, B. (2009) *Protein Sci.* **18**, 1241–1251
12. Yu, Z., Mirochnitchenko, O., Xu, C., Yoshizumi, A., Brodsky, B., and Inouye, M. (2010) *Protein Sci.* **19**, 775–785
13. Xu, C., Yu, Z., Inouye, M., Brodsky, B., and Mirochnitchenko, O. (2010) *Biomacromolecules* **11**, 348–356

Fusing Coiled-coil and Collagen Domains

14. Fields, G. B., and Noble, R. L. (1990) *Int. J. Pept. Protein Res.* **35**, 161–214
15. Qing, G., Ma, L. C., Khorchid, A., Swapna, G. V., Mal, T. K., Takayama, M. M., Xia, B., Phadtare, S., Ke, H., Acton, T., Montelione, G. T., Ikura, M., and Inouye, M. (2004) *Nat. Biotechnol.* **22**, 877–882
16. Demeler, B. (2005) *Modern Analytical Ultracentrifugation: Techniques and Methods* (Scott, D. J., Harding, S. E., and Rowe, A. J., Eds) pp. 210–229, Royal Society of Chemistry, London, United Kingdom
17. Hayes, D. B., Laue, T., and Philo, J. (1995–1998) *Sednterp*, Version 1.09, University of New Hampshire, Durham, NH
18. Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L., and Sonnhammer, E. L. (2000) *Nucleic Acids Res.* **28**, 263–266
19. Delorenzi, M., and Speed, T. (2002) *Bioinformatics* **18**, 617–625
20. Woolfson, D. N., and Alber, T. (1995) *Protein Sci.* **4**, 1596–1607
21. Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993) *Science* **262**, 1401–1407
22. Woolfson, D. N. (2005) *Adv. Protein Chem.* **70**, 79–112
23. Armstrong, C. T., Boyle, A. L., Bromley, E. H., Mahmoud, Z. N., Smith, L., Thomson, A. R., and Woolfson, D. N. (2009) *Faraday Discuss* **143**, 305–317
24. Feng, Y., Melacini, G., and Goodman, M. (1997) *Biochemistry* **36**, 8716–8724
25. Harrington, W. F., and Karr, G. M. (1970) *Biochemistry* **9**, 3725–3733
26. Engel, J., and Prockop, D. J. (1991) *Annu. Rev. Biophys. Biophys. Chem.* **20**, 137–152
27. Zitzewitz, J. A., Bilsel, O., Luo, J., Jones, B. E., and Matthews, C. R. (1995) *Biochemistry* **34**, 12812–12819
28. Steinmetz, M. O., Stock, A., Schulthess, T., Landwehr, R., Lustig, A., Faix, J., Gerisch, G., Aebi, U., and Kammerer, R. A. (1998) *EMBO J.* **17**, 1883–1891
29. Moutevelis, E., and Woolfson, D. N. (2009) *J. Mol. Biol.* **385**, 726–732

Designed Coiled Coils Promote Folding of a Recombinant Bacterial Collagen
Ayumi Yoshizumi, Jordan M. Fletcher, Zhuoxin Yu, Anton V. Persikov, Gail J. Bartlett,
Aimee L. Boyle, Thomas L. Vincent, Derek N. Woolfson and Barbara Brodsky

J. Biol. Chem. 2011, 286:17512-17520.

doi: 10.1074/jbc.M110.217364 originally published online March 28, 2011

Access the most updated version of this article at doi: [10.1074/jbc.M110.217364](https://doi.org/10.1074/jbc.M110.217364)

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

Supplemental material:

<http://www.jbc.org/content/suppl/2011/03/28/M110.217364.DC1>

This article cites 27 references, 8 of which can be accessed free at

<http://www.jbc.org/content/286/20/17512.full.html#ref-list-1>